Retention of Non-Traditional Adult Students: A Quantitative Approach

Presented at the 2008 AIR Annual Forum

Tracy Mohr, Senior Research Associate
Office of Enrollment Marketing and Research
May 2008

Contents

- Research Questions and Goals
- DePaul University and the School for New Learning
- The Fall 2000 Non-Traditional Student Cohort
- Using Survival Analysis and Logistic Regression to Analyze
- Concluding Thoughts

Research Question

☐How well can survival analysis help to understand attrition and retention patterns for non-traditional adult students, and what are these patterns?

Goals

- Investigate if a quantitative approach can be used to identify significant variables in an cohort of non-traditional undergraduate students.

- Determine if this approach can be used to possibly predict which non-traditional students may be at risk.

About DePaul University

- Founded in 1898 in Chicago, Illinois.

- 23,000 students in 2007.

- DePaul has 9 colleges and 8 campuses in Chicago and its suburbs.

- DePaul is the largest Catholic university in the US, and is the ninth largest private university.

- DePaul's enrollment grew by 71% from 1987 to 2007.

DePaul University's School for New Learning (SNL)

- Created in 1972 to serve the needs of adult students (age 24 and older).
- In Fall 2007, enrolled over 2,000 students.
- SNL gives students credit toward degrees for life experience. SNL undergraduate students design their own degree and take courses that meet their professional and personal goals.
- Courses are offered at all DePaul campuses and online. SNL students have the option of completing all coursework online.

Data Used in This Analysis

- For this analysis, the Fall 2000 SNL new student cohort (n=332) was used.

- For demographic comparison, a cohort of transfers (TRN) who were adult students (age 24 or higher) and entered colleges other than SNL in Fall 2000 was drawn. This yielded 213 students.

Selected Demographics

Fall 2000 SNL Cohort (n=332)

- 52% white, 30% African American, 11% Hispanic, 2% Asian
- 70% female
- 35% from Chicago, 54% from Chicago suburbs
- Median age at entry was 35; 57% between age 30-44, 20% age 45 or older
- 96% were enrolled part-time in their first quarter
- 28% received some type of financial aid

## Fall 2000 Adult TRN Cohort (n=213)

- 48% white, 15% African American, 12% Hispanic, 10% Asian
- 51% female
- 51% from Chicago, 36% from Chicago suburbs
- Median age at entry was 27; 27% between age 30-44, 3% age 45 or older
- 46% were enrolled part-time in their first quarter
- 77% received some type of financial aid

Retention and Attrition Issues

- **Key Points**
  - Students entering SNL were less than half as likely to graduate within six years as were adult transfer students.
  - SNL students were twice as likely not to return one year later (Fall 2001) as were adult transfer students.
  - 70% of SNL students left without graduating, compared to 40% of adult transfer students.

# 61% of all SNL students entering in Fall 2000 returned in Fall 2001, compared to 80% of adult transfer students.
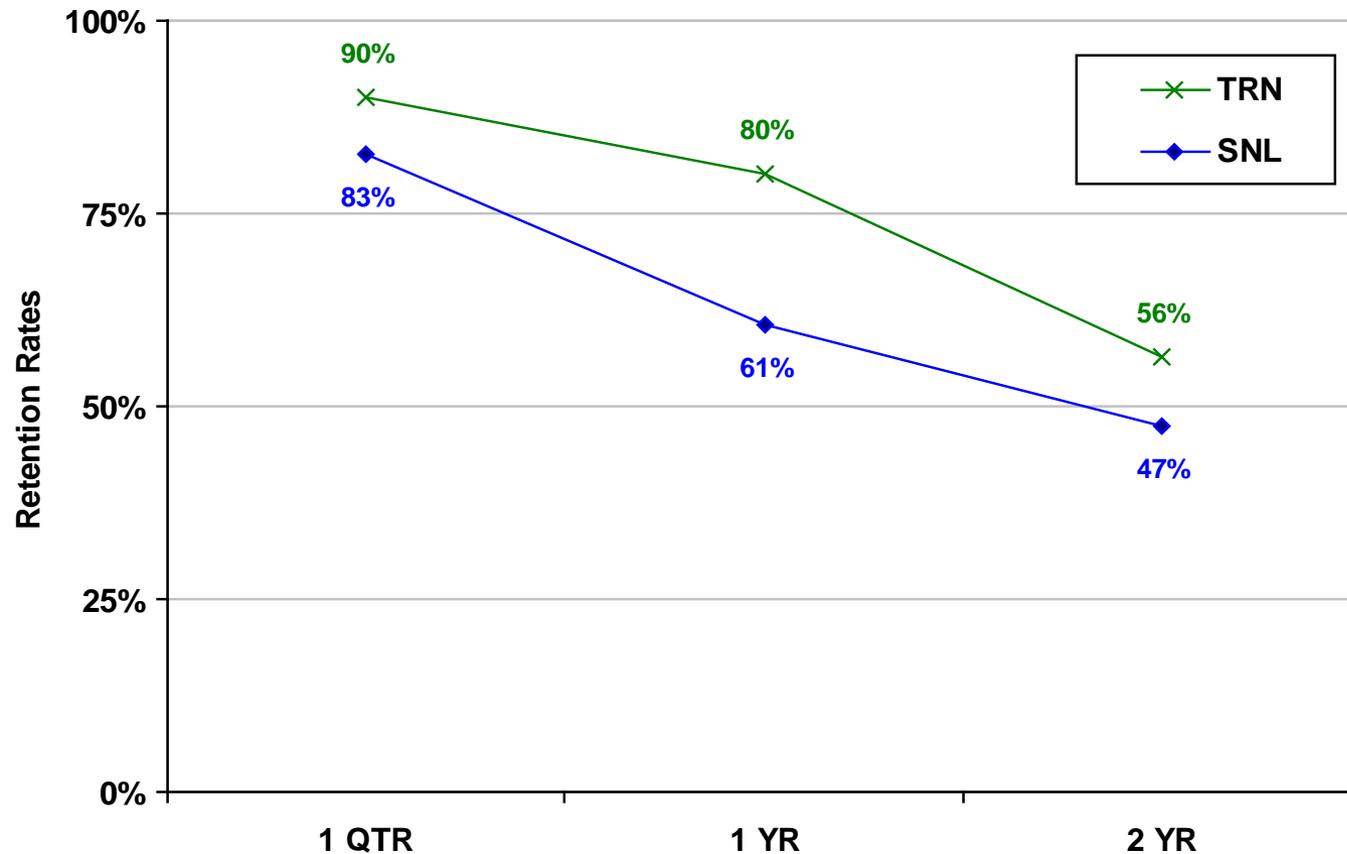
20% of all SNL students entering in Fall 2000 graduated within six years, compared to 53% of adult transfer students.
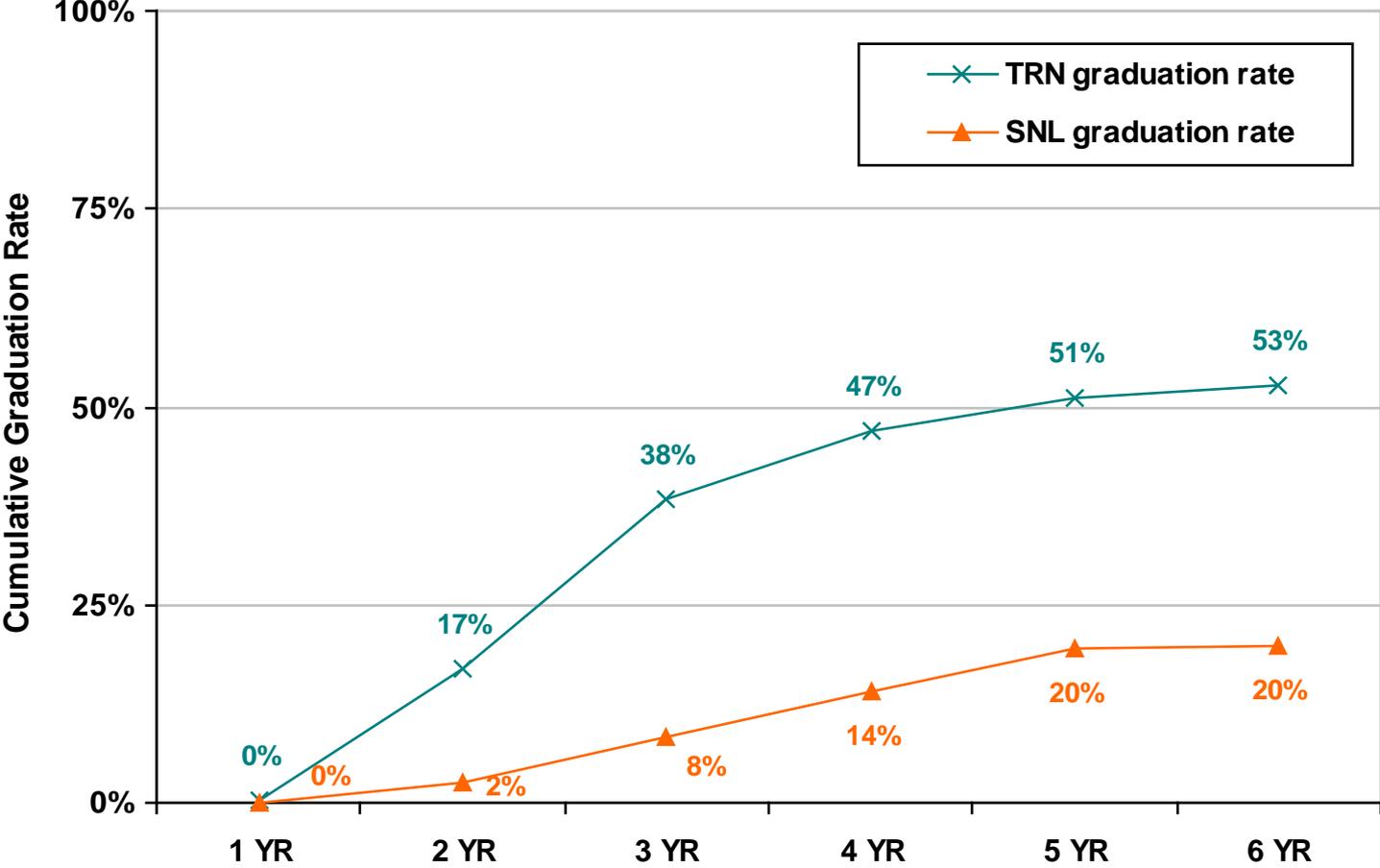
SNL students entering in Fall 2000 were almost twice as likely to leave within two years than were adult students entering as traditional transfers in Fall 2000.
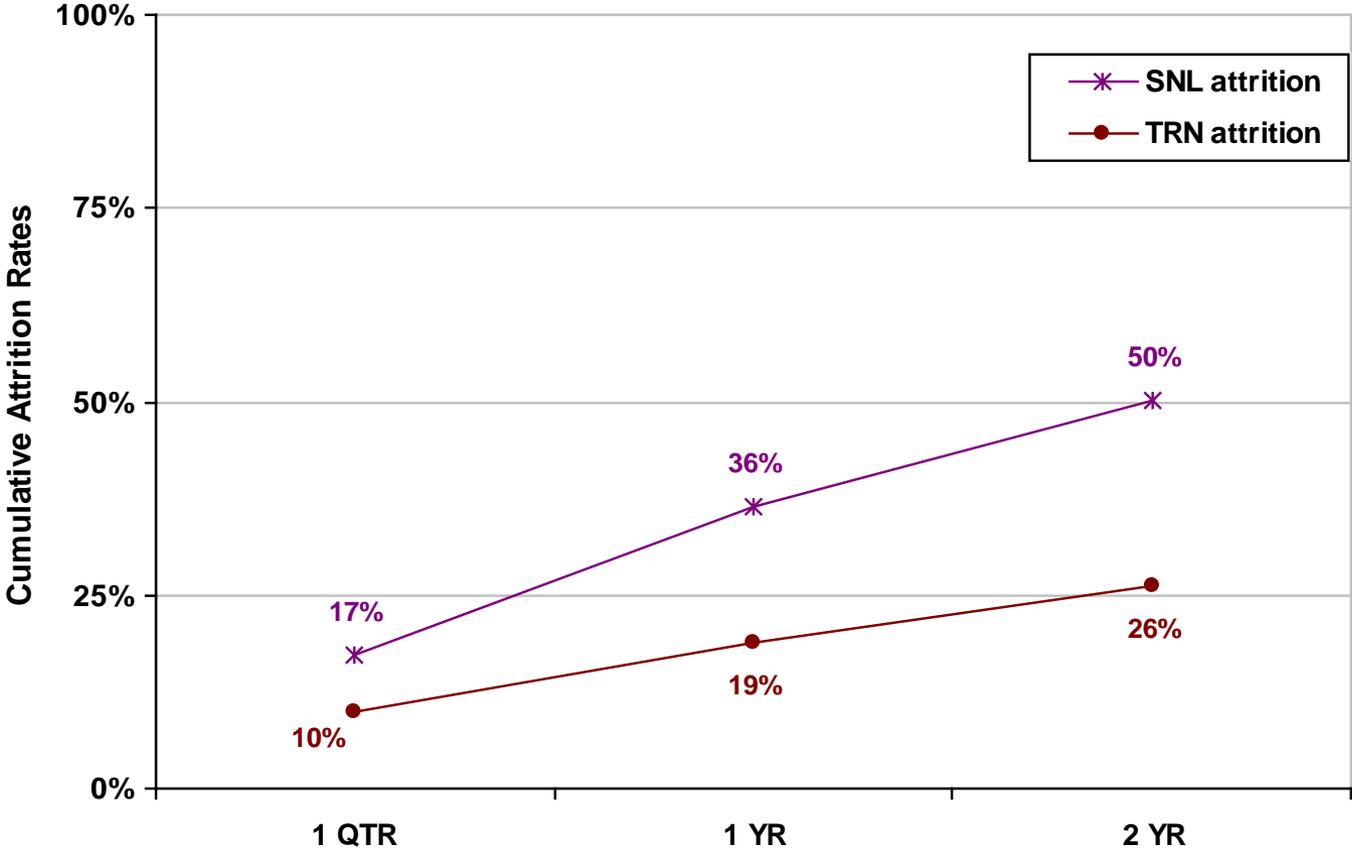
Survival Analysis Defined

- General definition: Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is *time until an event occurs.**

- Survival analysis techniques were primarily developed for use in medical and biological sciences (hence the name).

- Other names for survival analysis include event history analysis and reliability analysis.

*Kleinbaum and Klein, <u>Survival Analysis</u> (2005) (emphasis in original)

Survival Analysis Defined

- Censored data is at the heart of survival analysis.

- In general, censored observations arise whenever the dependent variable of interest represents the time to a terminal event, and the duration of the study is limited in time.

- Unlike other methods, survival analysis takes into account these 'incomplete' observations.

Steps Used in this Survival Analysis

- Survival Curves
- Survival Analysis Using Proportional Hazards (PH) Assumption
- Survival Analysis Using Accelerated Failure Time (AFT) Assumption

Survival Curves

- Illustrate differing survival behavior by student characteristic

- Variables analyzed:
  - Sex, Ethnicity, GPA group, Age group, Region, Financial Aid Recipient

Example – Female SNL students were more likely to persist than male students.

Survival Curves for Fall 2000 SNL Cohort
Adjusted Survival Curve Stratified by Sex

Survival Distribution Function Estimate

Number of Terms Enrolled

Gender   + + + F   + + + M

DE PAUL UNIVERSITY
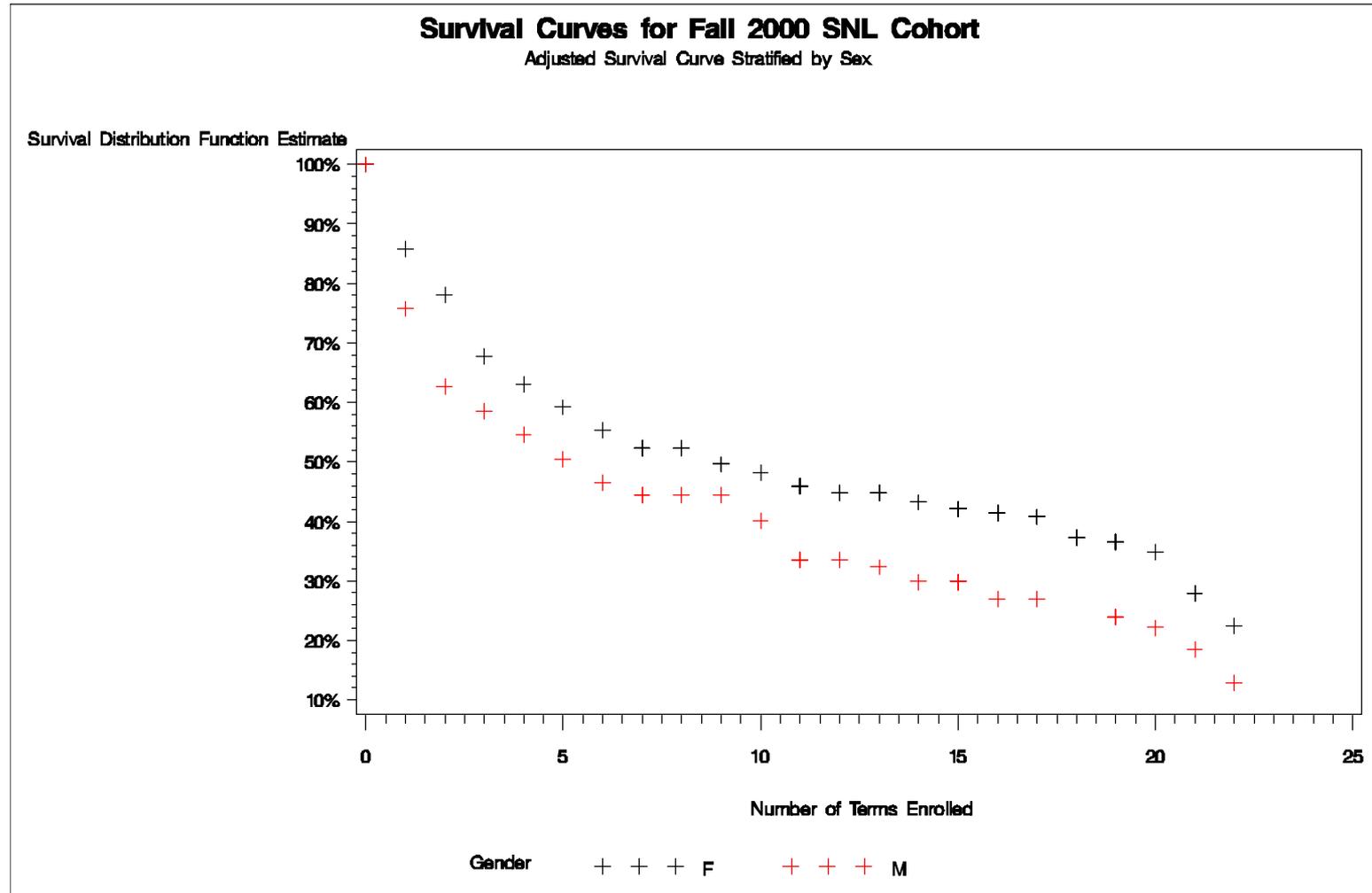
Example –SNL students receiving some type of financial aid were more likely to persist than those not receiving aid.

Survival Curves for Fall 2000 SNL Cohort
Adjusted Survival Curve Stratified By Aid Received
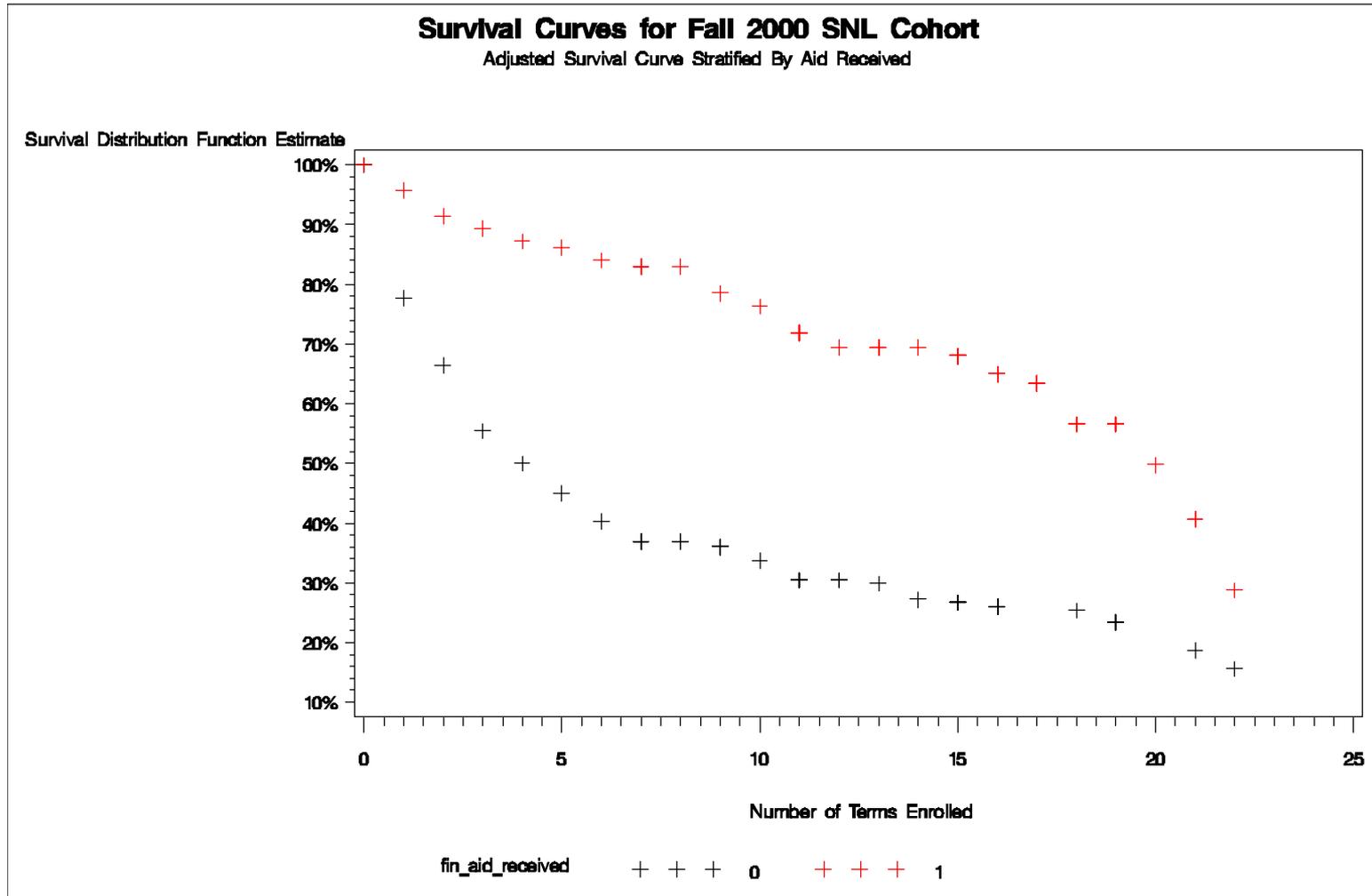
Application of Survival Analysis Models to SNL Cohort

- Time variable: dependent variable – in this study, number of quarters enrolled until event occurs.
- Censoring variable: a binary variable, perhaps better described as the event variable – in this case, student departing.
- Independent variables – must be numeric variables.
- Hazard: probability that the event will occur for an individual observation.

Proportional Hazards (PH) Model (also referred to as Cox Regression)

- Assumes that the hazard rate for any two observations is the same across time periods (proportionality assumption – refer to slide 16).

- Dependent variable is time variable.

- Independent variables are chosen – must be numeric variables.

- Hazard ratio – interpreted similarly to odds ratio in logistic regression.
  - In this model, a hazard ratio greater than 1 indicates a greater likelihood of departure.

Results of Survival Analysis on SNL Cohort Using PH Model

- **Using SAS proc PHREG:**

- Of fifteen independent variables, six were significant; three at $p < 0.05$, and three at $p < 0.15$.

- Variables significant at $p < 0.05$:
  - Student received student loans (binary: 1 = received loans)
  - First-quarter GPA
  - Age

- Variables significant at $p < 0.15$:
  - Student received other form of financial aid (not loans, Pell, or tuition reimbursement) (binary: 1 = received aid)
  - Gender (1 = female, 2 = male)
  - Student's permanent address was in Illinois, but not in Chicago or its suburbs (binary: 1 = lives elsewhere in Illinois)

- Using the TEST statement, it was found that the Proportionality Assumption was violated ($p < 0.01$). This test is based on a Wald chi-square statistic.

Significant Variables (p < 0.15) from Survival Analysis on SNL Cohort Using PH Model

Analysis of Maximum Likelihood Estimate							
Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
loans_received	-0.80496	0.26884	8.9650	0.0028	0.447	0.264	0.757
gpa1	-0.13353	0.04862	7.5436	0.0060	0.875	0.795	0.962
Age	-0.01742	0.00874	3.9756	0.0462	0.983	0.966	1.000
otheraid_received	-0.58363	0.35495	2.7035	0.1001	0.558	0.278	1.119
gender	0.24509	0.14932	2.6942	0.1007	1.278	0.954	1.712
region_otherIL	0.61724	0.40893	2.2783	0.1312	1.854	0.832	4.132

Violation of Proportionality

- **Violation of Proportionality: this indicates that a proportional hazard model may not be appropriate.**
  - Possible solutions include
    - adding time-dependent interaction variables (interactions of the predictors with the time variable),
    - stratifying on non-proportional predictors,
    - using a parametric regression model.

Results of Survival Analysis on SNL Cohort Using PH Model, Adding Time-Dependent Covariates to Account for Proportionality Violation

- **Using SAS proc PHREG:**

- Time-dependent covariates were created using the six variables that were significant in the preceding model.

- Of twenty-one independent variables, eight were significant at $p < 0.05$.

- Variables significant at $p < 0.05$:
  - Student received student loans (binary: 1 = received loans)
  - First-quarter GPA
  - Age
  - Time-dependent interaction: Student received student loans
  - Student received other form of financial aid (not loans, Pell, or tuition reimbursement) (binary: 1 = received aid)
  - Time-dependent interaction: Student received other form of financial aid
  - Time-dependent interaction: First-quarter GPA
  - Time-dependent interaction: Age

- *Proportionality Assumption was still violated, but incorporating the interaction effects serves to incorporate that non-proportionality*

Significant Variables (p < 0.15) from Survival Analysis on SNL Cohort Using PH Model, Adding Time-Dependent Covariates to Account for Proportionality Violation

Analysis of Maximum Likelihood Estimate							
Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
loans_received	-1.71199	0.49442	11.9897	0.0005	0.181	0.068	0.476
gpa1	-0.23020	0.07125	10.4399	0.0012	0.794	0.691	0.913
Age	-0.03320	0.01205	7.5904	0.0059	0.967	0.945	0.990
loan_time	0.09665	0.04008	5.8147	0.0159	1.101	1.018	1.191
otheraid_received	-1.25808	0.52214	5.8057	0.0160	0.284	0.102	0.791
otheraid_time	0.08526	0.03962	4.6305	0.0314	1.089	1.008	1.177
gpa1_time	0.01488	0.00696	4.5709	0.0325	1.015	1.001	1.029
age_time	0.00269	0.00126	4.5423	0.0331	1.003	1.000	1.005

Possible Issues Specific to This Survival Analysis

- **Certain types of financial aid may have restrictions:**
  - Aid may be contingent on full-time enrollment.
  - Aid may be capped or have a time limit.
- **Almost 75% of Fall 2000 SNL students had a GPA of 0.00 in their first quarter; the first course taken by most SNL students is pass/fail.**

Accelerated Failure Time (AFT) Model

- Parametric regression model (see previous slide).
- Assumes that survival time changes by a constant factor when comparing different levels of covariates.
- As with PH model:
  - Dependent variable is time variable.
  - Censoring variable is departure without a degree (1=departure, 0=censored).
  - Independent variables are chosen – must be numeric variables.

Results of Survival Analysis Using AFT Model

- **Using SAS proc LIFEREG:**

- Of fifteen independent variables, six were significant at $p < 0.15$.

- Variables significant at $p < 0.05$:
  - Student received student loans (binary: 1 = received loans), First-quarter GPA, Age

- Variables significant at $p < 0.15$:
  - Student's permanent address was in Illinois, but not in Chicago or its suburbs (binary), Student received other form of financial aid (not loans, Pell, or tuition reimbursement) (binary: 1 = received aid), Gender (1 = female, 2 = male)

- A positive Scale parameter tells us that the hazard is increasing.

- Hazard ratios (HR) for each covariate, controlling for the other covariates, are calculated as follows:
  - HR = exp[(Weibull Shape parameter)*(-(covariate)]
  - Example: Hazard Ratio for (loans_received = 1 vs. loans_received = 0)
    = exp[(0.9136)(-0.9746)] = 0.41

Significant Variables (p < 0.15) from Survival Analysis Using AFT Model

Analysis of Parameter Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	2.1079	0.8224	0.4959	3.7198	6.57	0.0104
loans_received	0.9746	0.2936	0.3992	1.5501	11.02	0.0009
gpa1	0.1509	0.0534	0.0463	0.2555	8.00	0.0047
Age	0.0213	0.0096	0.0025	0.0401	4.95	0.0262
region_otherIL	-0.8153	0.4457	-1.6888	0.0583	3.35	0.0674
otheraid_received	0.6817	0.3907	-0.0840	1.4474	3.04	0.0810
gender	-0.2485	0.1633	-0.5687	0.0716	2.32	0.1281
Scale	1.0946	0.0585	0.9857	1.2155		
Weibull Shape	0.9136	0.0489	0.8227	1.0145		

Modeling One-Year Retention of SNL Students

- As already noted, SNL students were more likely to leave after one year than adult transfers.

- Logistic regression using one-year retention as the dependent variable was used to look for significant variables.

  - Unlike survival analysis, logistic regression can use both numeric and character variables.

Results of Logistic Regression on One-Year Retention

- Using SAS proc LOGISTIC

- Of sixteen independent variables, five were significant at $p < 0.15$.

- Variables significant at $p < 0.05$:
  - Student received student loans in 2000-01 (binary)
  - Interaction effect of hours enrolled in each of first three quarters (Fall 2000, Winter 2001, Spring 2001)
  - Age

- Variables significant at $p < 0.15$:
  - Student's permanent address was in Illinois, but not in Chicago or its suburbs (binary)
  - Student received a Pell grant in 2000-01 (binary)

- R-squared = 0.1345

Significant Variables (p< 0.15) from Logistic Regression on One-Year Retention

Analysis of Maximum Likelihood Estimates					
Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Odds Ratio Point Estimate
Intercept	-0.7664	1.6814	0.2078	0.6485	
loans_received01	1.7806	0.7348	5.8717	0.0154	5.933
hours1*hours2*hours3	0.0020	0.0009	4.6832	0.0305	
Age	0.0335	0.0156	4.5842	0.0323	1.034
region_otherIL	-1.4427	0.7563	3.6394	0.0564	0.236
pell_received01	-1.4874	0.8953	2.7600	0.0966	0.226

Conclusions

- It is evident that non-traditional adult students present different retention issues than traditional adult transfer students.

- In all models, financial aid status (particularly student loan status), first-term GPA, and age were significant variables.

- Because it uses both time-dependent and censored data, survival analysis offers possible advantages over logistic regression.

- Given the large percentage of non-traditional students taking a pass-fail course in their first quarter, first-quarter GPA, traditionally an excellent indicator of retention, may present unique issues.

Actions and Recommendations

- Review results with decision makers in admissions

- Examine SNL cohorts from winter and spring quarters.

- Analyze adult transfer cohort for comparison.

- Test model on subsequent cohorts for possible predictive value.

Appendix

- Complete results of models.

Survival Analysis on SNL Cohort Using PH Model

Analysis of Maximum Likelihood Estimate								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
loans_received	1	-0.80496	0.26884	8.9650	**0.0028**	0.447	0.264	0.757
gpa1	1	-0.13353	0.04862	7.5436	**0.0060**	0.875	0.795	0.962
Age	1	-0.01742	0.00874	3.9756	**0.0462**	0.983	0.966	1.000
otheraid_received	1	-0.58363	0.35495	2.7035	**0.1001**	0.558	0.278	1.119
sex	1	0.24509	0.14932	2.6942	**0.1007**	1.278	0.954	1.712
region_otherIL	1	0.61724	0.40893	2.2783	**0.1312**	1.854	0.832	4.132
pell_received	1	0.27614	0.36580	0.5699	0.4503	1.318	0.644	2.700
region_suburb	1	0.15603	0.26958	0.3350	0.5627	1.169	0.689	1.983
race_hisp	1	-0.15250	0.34863	0.1914	0.6618	0.859	0.434	1.700
race_black	1	0.12733	0.29561	0.1855	0.6667	1.136	0.636	2.027
CountOfMajor	1	0.16957	0.46225	0.1346	0.7137	1.185	0.479	2.932
race_white	1	-0.09360	0.28256	0.1097	0.7404	0.911	0.523	1.584
reimb_received	1	-0.08170	0.35104	0.0542	0.8160	0.922	0.463	1.834
race_asian	1	-0.09815	0.49276	0.0397	0.8421	0.907	0.345	2.381
region_chi	1	0.05122	0.27972	0.0335	0.8547	1.053	0.608	1.821

Survival Analysis on SNL Cohort Using AFT Model

Analysis of Parameter Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	2.1079	0.8224	0.4959	3.7198	6.57	**0.0104**
loans_received	0.9746	0.2936	0.3992	1.5501	11.02	**0.0009**
gpa1	0.1509	0.0534	0.0463	0.2555	8.00	**0.0047**
Age	0.0213	0.0096	0.0025	0.0401	4.95	**0.0262**
region_otherIL	-0.8153	0.4457	-1.6888	0.0583	3.35	**0.0674**
otheraid_received	0.6817	0.3907	-0.0840	1.4474	3.04	**0.0810**
sex	-0.2485	0.1633	-0.5687	0.0716	2.32	**0.1281**
race_black	-0.2568	0.3226	-0.8890	0.3755	0.63	0.4261
pell_received	-0.3007	0.4045	-1.0935	0.4921	0.55	0.4572
region_suburb	-0.1996	0.2949	-0.7775	0.3783	0.46	0.4985
region_chi	-0.1137	0.3055	-0.7125	0.4851	0.14	0.7098
CountOfMajor	-0.1780	0.5048	-1.1675	0.8114	0.12	0.7244
race_hisp	0.1087	0.3814	-0.6388	0.8562	0.08	0.7756
reimb_received	0.0943	0.3843	-0.6588	0.8474	0.06	0.8061
race_asian	0.1227	0.5374	-0.9305	1.1759	0.05	0.8194
race_white	0.0110	0.3089	-0.5944	0.6165	0.00	0.9715
Scale	1.0946	0.0585	0.9857	1.2155		
Weibull Shape	0.9136	0.0489	0.8227	1.0145		

Logistic Regression Results

Analysis of Maximum Likelihood Estimates					
Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Odds Ratio Point Estimate
Intercept	-0.7664	1.6814	0.2078	0.6485	
loans_received01	1.7806	0.7348	5.8717	**0.0154**	5.933
hours1*hours2*hours3	0.00204	0.000943	4.6832	**0.0305**	
Age	0.0335	0.0156	4.5842	**0.0323**	1.034
region_otherIL	-1.4427	0.7563	3.6394	**0.0564**	0.236
pell_received01	-1.4874	0.8953	2.7600	**0.0966**	0.226
sex	-0.3270	0.2707	1.4590	0.2271	0.721
reimb_received01	0.8306	0.7042	1.3911	0.2382	2.295
region_suburb	-0.5209	0.5035	1.0701	0.3009	0.594
gpa1	0.0971	0.0949	1.0457	0.3065	1.102
otheraid_received01	0.7721	0.7652	1.0180	0.3130	2.164
CountOfMajor	0.9067	1.1678	0.6029	0.4375	2.476
race_black	-0.6345	0.8402	0.5703	0.4501	0.530
region_chi	-0.3558	0.5254	0.4587	0.4982	0.701
race_white	-0.5224	0.8151	0.4108	0.5216	0.593
race_other	0.5709	1.0413	0.3006	0.5835	1.770
race_hisp	-0.3104	0.8920	0.1211	0.7278	0.733